Sharing Guilt:

How Better Access to Information May Backfire*

Roman Inderst

Goethe University Frankfurt, Theodor-W.-Adorno-Platz 1, D-60323 Frankfurt am Main, Germany,
(e-mail: inderst@finance.uni-frankfurt.de)

Kiryl Khalmetski

University of Cologne, Albertus-Magnus-Platz, D-50923 Cologne, Germany
(e-mail: kiryl.khalmetski@uni-koeln.de)

Axel Ockenfels

University of Cologne, Albertus-Magnus-Platz, D-50923 Cologne, Germany
(e-mail: ockenfels@uni-koeln.de)

2 April 2018

Abstract

We study strategic communication between a customer and an advisor who is privately informed about the most suitable choice for the customer but whose preferences are misaligned with the customer's preferences. The advisor sends a message to the customer who, in turn, can secure herself from bad advice by acquiring costly information on her own. In our experiments, we find that making the customer's information acquisition less costly leads to *less* prosocial behavior of the advisor. This can be explained by a model of shared guilt, which predicts a shift in causal attribution of guilt from the advisor to the customer if the latter could have avoided her *ex post* disappointment. We conclude that providing better access to information through, e.g., consumer protection regulation or digital information aggregation and dissemination, may have unintended negative consequences on peoples' willingness to take responsibility for each other.

Keywords: shared guilt, trust, guilt aversion, responsibility diffusion, advice

JEL classification: C91, D82, D83

* Accepted for publication in *Management Science.*

1. Introduction

Informational transparency is often viewed as a fundamental prerequisite for efficient trading and the smooth functioning of markets and institutions (Akerlof 1970, Stiglitz and Weiss 1981, Kaufmann and Bellver 2005). For instance, much effort in consumer protection regulation has been put into making it easier for consumers to directly assess and compare the various products and options that are available to them, rather than having to rely on the advice of sellers or brokers, who often have conflicting interests. Examples of this approach abound and include the standardization of product information, the presentation of information in transparent language, or the use of labels (Golan et al. 2001, Kozup and Hogarth 2008). Similarly, recent advances in computer technology have helped to substantially reduce consumers' uncertainty about product quality, sellers' trustworthiness, and competitors' price offers in digital markets and e-commerce (Bakos 1997, Goldmanis et al. 2010, Bolton et al. 2013). Such regulation and digitalization of economic interactions are supposed to help customers make better decisions.

In this paper, however, we show that, if social preferences come into the picture, informational transparency can backfire. To the extent regulators make it easier for consumers to protect themselves against bad decisions, professional advisors may feel less socially responsible to steer them away from an unwise choice. We provide controlled laboratory evidence that a policy that facilitates information transparency may crowd out social responsibility of advisors for the consumer. We then propose a model that describes the underlying behavioral mechanism.[1]

In particular, we consider a communication game where an uninformed customer can consult with an informed advisor before making a product choice. If the customer consults with the advisor, the latter has a strict monetary incentive to lie to the customer. At the same time, the customer has an option to acquire information on her own at a fixed cost. We find that the immediate benefits from easing access to product information, by lowering the costs to customers of getting fully informed by themselves, may be overcompensated. The reason is that when customers have a lower cost of obtaining information but still rely on the advisor, they are more likely to receive unsuitable advice than when obtaining that outside information is more expensive. That is, if customers forgo the opportunity to inform themselves at relatively low costs, advisors are less willing to act in the customers' interest, and rather lie, even though they rightly

[1] This countervailing effect of an informational transparency policy is difficult to identify in a naturally occurring field setting. The only exception we are aware of is Ahmad et al. (2006), who show that physicians may become less helpful to a patient once they learn that he or she is using the Internet to access health information from different sources. While we focus on cases where customers can make good use of an informational transparency policy, such policy has also been criticized in instances where consumers are expected to make little use of this additional information, e.g., due to information overload (Lacko and Pappalardo 2007). Other contributions focus on greater disclosure of conflicts of interest, arising for instance from commissions, and on how this affects firms' strategic behavior (e.g., Inderst and Ottaviani 2012, Anagol and Kim 2012, Duarte and Hastings 2012).

anticipate that customers who self-select into seeking advice put more trust in them. Moreover, customers tend to not anticipate that advisors will be less trustworthy if the cost of information decreases. In particular, those customers who refer to the advisor generally anticipate precisely the opposite tendency, further reinforcing the potential downside of a mandated transparency policy.

On a conceptual level, we show that the observed behavior of advisors can be explained by a natural extension of the concept of guilt in psychological game theory. Psychological game theory (Geanakoplos et al. 1989, Battigalli and Dufwenberg 2009) postulates that individual preferences accommodate not only monetary payoffs, but also beliefs, including beliefs about others' beliefs. This way, it can capture that behavior depends on foregone outside options. However, applying the canonical model of guilt aversion (Battigalli and Dufwenberg 2007, 2009) to our context would suggest that a better foregone outside option of the customer would unambiguously lead to *less* lying – the opposite of our finding. The reason is that the customer's choice to forego this option reveals her higher expectation of the advisor's trustworthiness, who in turn would be less inclined to lie because this would yield greater customer disappointment and thus generate more guilt according to the model. We extend this model to allow for a shift of causal attribution of guilt from the advisor to the customer if the latter could have avoided the loss. The underlying idea is that guilt is *shared* by the players according to the extent each player can be blamed for the outcome. We show how shared guilt allows for an inverse relationship between the customer's cost to choose the (foregone) outside option and the advisor's tendency to lie. There are two conflicting forces at work when information could have been obtained at a smaller cost, yet the customer nevertheless relies on the advisor: (*i*) more revealed trust in the advisor and thus more guilt from lying, and (*ii*) a shift in the attribution of guilt to the customer. Our main experimental setting allows us to disentangle these two conflicting forces and indeed finds evidence for both. However, the second effect dominates, which implies an overall negative relation between the cost of information for the customer and the advisor's likelihood of lying.

That being said, we emphasize that other psychological mechanisms not built on guilt aversion might contribute to our observations. For example, the choice of the customer to avoid paying a very low price for information may be interpreted by the advisor in a way unfavorable to the customer, e.g., as revealing her greediness, laziness, or specific preference characteristics like low loss aversion. This, in turn, may affect the advisor's preferences with respect to such customer.[2] Our second experiment is aimed to rule out such explanations by making these signaling effects orthogonal to the current interaction with the advisor. At the same time, our mathematical model (in particular, its generalized version in Appendix A) can be reinterpreted to reflect this or other alternative explanations that lead to a reduction in the advisor's

[2] Levine (1998) and Ellingsen and Johannesson (2008) propose models where individual preferences depend on the perceived opponent's type as revealed by the latter's actions.

perceived responsibility towards the customer due to the latter's reluctance to acquire information on her own.

Our main hypothesis relates to what Charness and Rabin (2002) called the "complicity effect," which posits that the mere fact that a trustor is an active player might diffuse the trustee's responsibility for the final outcome. Indeed, in their experiment they observed no evidence that a trusting choice by the trustor leads to more prosocial behavior of the trustee.[3] These results are consistent with our findings and model, yet an important difference of our approach is that we observe the effect of responsibility diffusion by varying *only* the outside option of the customer (i.e., the trustor) while keeping her actual choice whether to trust fixed, thus controlling for an important element of positive reciprocity.[4]

More generally, our results align with earlier findings of Charness (2000), who asserted and experimentally showed that individuals behave more prosocially in situations where they have full responsibility for the outcome, and correspondingly less so if this responsibility can be shared with another party.[5] This effect was further confirmed in studies showing that responsibility can be diffused among several players who jointly cause harm to another player through delegation of decisions (Fershtman and Gneezy 2001, Hamman et al. 2010, Coffman 2011, Bartling and Fischbacher 2012, Oexl and Grossman 2013), or group decision making (Falk and Szech 2013, Bartling et al. 2015).[6] Unlike the approaches taken in this literature, ours involves the negatively affected player in the sharing of responsibility.[7]

Our model and experiments reconcile seemingly inconsistent results in the experimental literature on belief-dependent preferences. On the one hand, various studies showed that higher second-order beliefs *all*

[3] Brandts et al. (2015) observed no evidence for reciprocating trust in a similar experiment. Cappelen et al. (2016) showed that individuals tend to consider another subject as deserving a low *ex post* outcome in the case that the latter was making an explicit choice, even if there was neither a plausible alternative, nor a causal link between the choice and the outcome.

[4] The experiment of Dufwenberg and Gneezy (2000) was also based on varying the outside option of a trustor in a variant of the trust game (the Lost Wallet Game). At the same time, the payoff structure in their experiment implied that the trusting choice of the trustor was explicitly kind to the trustee (as otherwise he got zero payoff). This allowed for a larger scope for positive reciprocity than our experiment, where the payoff structure was adjusted to disentangle the effect of shared guilt.

[5] Charness (2000) termed this as "the responsibility-alleviation effect". Charness et al. (2012) further showed that subjects might even *strategically* delegate the full responsibility over outcome to another party in order to trigger a prosocial response (i.e., by strategically reducing the alleviation of responsibility).

[6] Garofalo and Rott (forthcoming) showed that decision makers may sometimes attempt to shift blame by the delegation of *communicating* an unfair allocation (unlike delegation of the decision rights as in the aforementioned literature). Interestingly, such attempts are generally not successful as they do not lead to less punishment of the decision maker by the negatively affected players, and can even backfire.

[7] Bartling and Fischbacher (2012) also propose a model of diffusion of responsibility between active players whose decisions affect another party based on a belief-based measure of anticipated responsibility, which bears similarities to the way we model shared-guilt preferences. Yet, Bartling and Fischbacher (2012) do not consider a dependence of responsibility attribution on second-order beliefs, which is at the core of our model. A related model of Engl (2017) assigns responsibility based on distance to being pivotal for the outcome. In contrast to our approach, the model of Engl (2017) does not allow for the responsibility to be diffused among two players if both of them are pivotal for the outcome.

else being equal lead to more trustworthiness (e.g., Charness and Dufwenberg 2006, Reuben et al. 2009, Khalmetski et al. 2015, Khalmetski 2016).[8] This is consistent with both the standard model of guilt aversion and our model of shared guilt, as well as with our experimental results. On the other hand, multiple experiments have shown that more trusting *behavior*, potentially signaling higher expectations, does not lead to more trustworthiness (Dufwenberg and Gneezy 2000, Servátka & Vadovič 2009, Cox et al. 2010, Beck et al. 2013, Woods and Servátka 2016). This is inconsistent with what is suggested by the standard model of guilt aversion, yet also consistent with our extended model and our experimental results.

Our work also contributes to the experimental literature (which starts with Gneezy 2005) on the determinants of subjects' tendency to lie. One of the key findings in this literature is that lying is affected not only by the monetary consequences for the liar, but also by the consequences for the receiver of the message (Gneezy 2005, Erat and Gneezy 2012). Our experiments add that not only materialized consequences but also unrealized consequences (outside options) of the receiver affect the propensity to lie in systematic ways.

Our results show that increased informational transparency can make customers who still prefer to trust in others' advice more vulnerable. As far as advisors' conduct is not completely constrained by rules and liability concerns, their advice to customers may become more biased, because of a shift of blame that crowds out advisors' social concerns for customers. While a policy implemented to ease access to information may thus have the intended consequences for those customers who in fact sidestep advice, it may backfire for those customers who still choose to rely on their possibly conflicted advisors.[9]

Section 2 presents our theoretical setting. Section 3 describes the baseline experiment. Section 4 provides the results of a laboratory robustness check, and Section 5 concludes.

2. Modeling shared guilt

2.1. Monetary payoffs in the Sharing Guilt Game

Suppose two players, an advisor (he) and a customer (she), play the Sharing Guilt Game shown in Fig. 1. In the first stage, the customer chooses between *In* and *Out*. In case of *Out*, the game ends and the customer gets her outside option $\gamma \in (0,1)$, while the advisor gets a low payoff π_a^L. In case of *In*, the game proceeds to the next stage where the advisor chooses between two actions: *Truth* or *Lie*. In the final stage,

[8] See also Ellingsen et al. (2010) and Vanberg (2008) for conflicting evidence; see Khalmetski et al. (2015) for a discussion. Interestingly, the dictator game in Vanberg (2008) involves a third party that allows responsibility diffusion in the spirit of our model; see Kawagoe and Narita (2014) for a discussion.

[9] A different backfiring mechanism through less suitable advice has been recognized by Cain et al. (2005, 2011): They show that disclosing a conflict of interest between advisor and advisee can induce less trustworthy behavior of advisors.

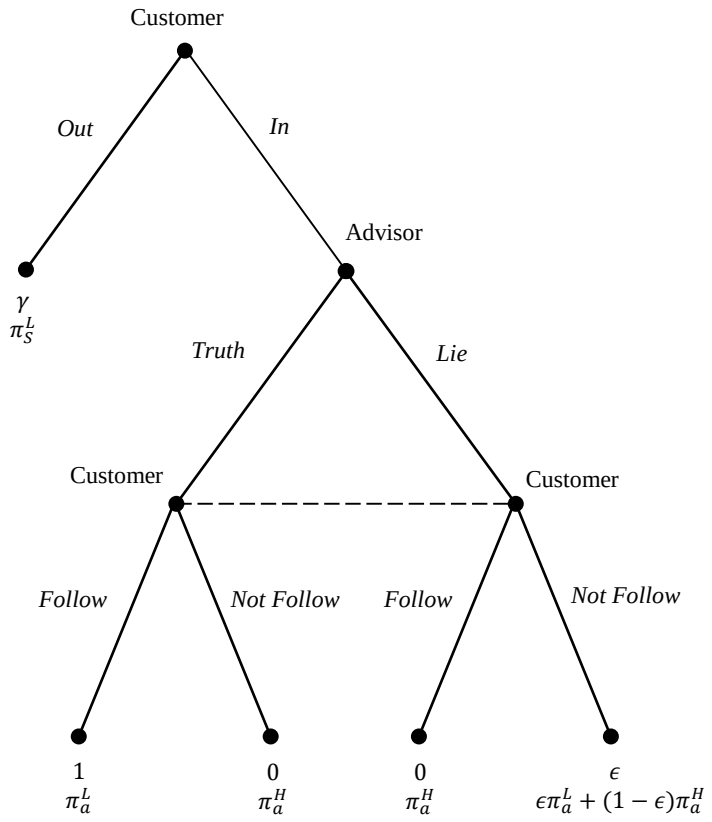

Fig. 1. Monetary Payoffs in the Sharing Guilt Game.

the customer chooses between *Follow* or *Not Follow*, not knowing whether the advisor lied or not. If the advisor tells the truth, the customer gets a payoff normalized to 1 if she follows the advisor, and a payoff of 0 if she does not follow him. If the advisor lies, the customer gets 0 if she follows the advisor. If she does not follow him in this case, she gets 1 with a small probability ϵ, and 0 with the remaining probability. One interpretation is that in this case the customer relies on her own limited knowledge in taking some final decision, or she just makes some random decision and hopes for good luck, so expected payoffs are small. The advisor earns a high monetary payoff $\pi_a^H > \pi_a^L$ if and only if the customer gets 0. With these monetary payoffs and sufficiently small ϵ, in a subgame perfect equilibrium, the customer plays *Out*.[10]

The game reflects situations in which players have opposing interests, and thus trust and trustworthiness cannot occur in equilibrium when all that matters is monetary payoffs. One application is professional (e.g., financial, retail or medical) advice about the most suitable option for a customer. The advisor can either

[10] The customer can earn a sufficiently high payoff after *In* only if she subsequently plays *Follow* with a non-negligible probability. However, then the advisor will always lie, which leads to zero payoff for the customer. Hence, for sufficiently small ϵ the customer's expected payoff from *In* in equilibrium is always smaller than the outside option γ.

recommend the right option, or mislead the customer for some monetary benefit, such as commissions from product providers when the customer follows bad advice.[11] At the same time, the customer can choose an outside option that yields the same payoff as the informed decision yet without referring to the advisor. However, the outside option only comes at positive cost since $\gamma < 1$. This can be interpreted as a costly effort to learn to become informed herself.

2.2. Psychological payoffs and predictions

We now allow that the advisor is not exclusively motivated by monetary payoffs, but may also be driven by social concerns that may result in a positive probability of truth-telling (as considered below). Then, while under sufficiently small ϵ, the customer's strategy (*In, Not Follow*) is still dominated by *Out*, a risk-neutral payoff maximizer prefers the strategy (*In, Follow*) over the outside option if

$$\alpha_c \geq \gamma, \tag{1}$$

where α_c is the probability of *Truth* expected by the customer.

Our main question is how the advisor's behavior conditional on *In* depends on the value γ of the customer's foregone outside option. Outcome-based models of social concerns (e.g., Bolton and Ockenfels 2000, Fehr and Schmidt 1999) would not predict any such effect since the mapping of the monetary payoffs of *both* players to the advisor's possible actions do not depend on γ once the customer chooses *In*. However, if the advisor's preferences incorporate belief-dependent payoffs, such effects become possible.

2.2.1. Simple guilt

According to the general model of guilt aversion (Battigalli and Dufwenberg 2007), the psychological cost inflicted on the customer in our game is given by the degree D_c to which her *ex ante* expectations are disappointed:

$$D_c(s_c, s_a) = \max\{0, E_c^0 - \pi_c(s_c, s_a)\}. \tag{2}$$

Here, π_c is the realized monetary payoff of the customer as a function of s_c and s_a, the strategies of the customer and the advisor, respectively, and E_c^0 is the customer's *ex ante* payoff expectation. The advisor's *simple guilt* is then determined as the share of D_c that could have been avoided by the advisor (i.e., which can be attributed to the advisor's choice):

[11] An advisor's concern for customers may also derive from liability and threat of prosecution. Still, in many instances, unsuitable advice may not be easily detectable and the likelihood of prosecution may remain sufficiently low to act as a sufficient deterrent. It is precisely in such situations where the danger is highest that informational transparency policies could have unwanted consequences by "crowding out" advisors' guilt.

$$G_a(s_c, s_a) = D_c(s_c, s_a) - \min_{\widetilde{s_a}} D_c(s_c, \widetilde{s_a}), \tag{3}$$

where the minimum is taken over the advisor's available strategies. The advisor's utility is given by

$$U_a(s_c, s_a) = \pi_a(s_c, s_a) - \theta_a G_a(s_c, s_a), \tag{4}$$

where θ_a is the advisor's individual sensitivity to simple guilt.

Conditional on *In*, the advisor can fully avoid letting down the customer by playing *Truth*, so that $\min_{\widetilde{s_a}} D_c(s_c, \widetilde{s_a}) = 0$. If the advisor chooses instead *Lie*, his simple guilt is

$$G_a(In, Lie) = D_c(In, Lie) - 0 = E_c^0 - \pi_c(In, Lie) = \alpha_c, \tag{5}$$

where we used that the customer earns 0 if she follows *Lie* and 1 if she follows *Truth*.

We assume that the customer's beliefs are heterogeneous in the population and *ex ante* unknown to the advisor (see, e.g., Bellemare et al. 2011, for experimental evidence, and Khalmetski et al. 2015 and Attanasi et al. 2015 for models based on a similar assumption). Such heterogeneity may arise, for instance, from customers' differences in beliefs about the advisor's guilt aversion, which may in turn depend on differences in their past experiences. Moreover, while our model is consistent with full customer rationality, we do not require that individual beliefs of customers correspond to a rational expectations equilibrium in the sense that they are fully consistent with the advisor's equilibrium strategies and beliefs. Indeed, it appears plausible that customers are unaware not only of the actual distribution of the advisors' preference parameters, but also of the distribution of beliefs of other customers in the population, as well as about the distribution of the advisors' second-order beliefs about these beliefs and so on.[12]

Given that the advisor's utility when $s_c = In$ and $s_a = Lie$ in (4) is linear in $G_a(In, Lie)$ and thus from (5) linear in α_c, to derive the advisor's expected utility we need only the first moment of his second-order beliefs, which we denote by β_a^I. That is, β_a^I denotes the advisor's second-order belief about α_c conditional on observing *In*:

$$\beta_a^I := E_a[\alpha_c | In] = E_a[\alpha_c | \alpha_c \geq \gamma]. \tag{6}$$

This reflects that a customer will only choose *In* when her belief about the advisor's trustworthiness satisfies the incentive constraint $\alpha_c \geq \gamma$. The advisor's expected utility from choosing *Lie* after *In* is then

$$E_a[U_a(In, Lie)] = \pi_a^H - \theta_a \beta_a^I. \tag{7}$$

As the advisor expects π_a^L from choosing *Truth*, together with (7) we thus have that he prefers to lie if and only if

[12] Attanasi et al. (2015) develop an approach to characterize psychological equilibrium under incomplete information about players' heterogeneous beliefs.

$$\pi_a^H - \theta_a \beta_a^I \geq \pi_a^L. \quad (8)$$

As far as $\beta_a^I > 0$ (which is implied by (6)), by (8) the advisor prefers lying if and only if

$$\theta_a \leq \theta_a^* = \frac{\pi_a^H - \pi_a^L}{\beta_a^I}. \quad (9)$$

Thus, for given $\beta_a^I > 0$, the advisor's optimal strategy can be characterized by the indifferent cutoff type θ_a^* such that only those types whose individual guilt sensitivity exceeds the cutoff tell the truth. Hence, the theory of simple guilt can explain why individual advisors do not lie and, as exhibited in equation (8), relates the advisor's decision to his second-order belief β_a^I about the customer's expectation of the advisor's probability to choose *Truth*.

Consider now the effect of the outside option γ on the advisor's decision to lie. With simple guilt, the size of γ affects the advisor's decision whether to choose *Truth* or *Lie* when it affects the advisor's second-order belief β_a^I. Once the customer's first-order beliefs are sufficiently dispersed in the population, (6) implies that the advisor's conditional expectation about the customer's beliefs β_a^I is strictly increasing in γ. Consequently, given (9), the cut-off level θ_a^*, and hence the rate of lying, becomes smaller as γ increases: As the choice of *In* reveals a higher level of trust, this leads to more trustworthiness of a guilt-averse advisor. This is summarized in the following prediction.

Prediction 1 [Simple Guilt]: *Under simple guilt, the advisor becomes less likely to lie (i.e., he tells the truth also for a lower level of guilt sensitivity) as the customer's payoff γ from the foregone outside option increases.*

2.2.2. Shared guilt

Shared guilt is a new concept that introduces a refined perspective on the attribution of blame and guilt. The idea is that if trust is abused, both the advisor's and the customer's actions in our Sharing Guilt Game can be considered as causes for the ultimate disappointment of the customer's initial expectations, which in turn diffuses responsibility for the outcome. In particular, both the advisor's and the customer's choice are *pivotal* for the customer's resulting losses: the outcome would not have emerged had one of the players deviated from his or her choice. Hence, as suggested by research in psychology and economics, the perceived responsibility for the outcome (in our case, the customer's disappointment) will be split between the two causing players.

Empirical support in psychology comes from research showing that the feeling of self-blame depends on how much a person perceives that she could have avoided a negative outcome had she acted differently (Miller and Turnbull 1990, Davis et al. 1996, Mandel 2003). Moreover, according to these studies, increased

self-blame, resulting from higher perceived avoidability of the outcome, leads to lower attribution of responsibility to another party, whose actions also contributed to the outcome and who is subsequently less blamed. In turn, if the latter party anticipates that his assigned responsibility will be alleviated, his psychological costs of feeling guilty for own action can then be naturally assumed to decrease, too (as is asserted in our further analysis). Regarding pivotality, in an overview of the relevant psychological literature, Alicke et al. (2015) demonstrate a strong link between being pivotal for outcomes and one's perception of moral responsibility. In the economics literature, Bartling et al. (2015) provide recent experimental evidence that the degree of pivotality of one's actions affects subsequent blame assignment.

In this section, we introduce a model of shared guilt, which captures the idea that the attribution of guilt for disappointing trust is *shared* between players whose choices eventually caused this disappointment, including the disappointed player herself. We do so in the most parsimonious and straightforward way. Formally, the psychological cost of the customer from being let down, $D_c(s_c, s_a)$, is a function of *both* the advisor's and the customer's strategy. Hence, in the same way the advisor is treated in (3), we can derive the part of the customer's disappointment that can be causally attributed to her own behavior, i.e., which could have been avoided had she deviated from her initial plans:

$$G_c(s_c, s_a) = D_c(s_c, s_a) - \min_{\tilde{s}_c} D_c(\tilde{s}_c, s_a). \tag{10}$$

We refer to this as "*self-blame*" of the customer. When the customer chooses *In* and the advisor responds with *Lie*, recall from (2) that the respective degree of letting down is given by $D_c(s_c, s_a) = E_c^0 - \pi_c(In, Lie) = \alpha_c$. Now with (10), the customer's self-blame is

$$\begin{aligned} G_c(In, Lie) &= \alpha_c - \min_{\tilde{s}_c}(\alpha_c - \pi_c(\tilde{s}_c, Lie)) \\ &= \alpha_c - (\alpha_c - \gamma) = \gamma, \end{aligned} \tag{11}$$

given that the customer could have avoided being let down by choosing *Out* instead of *In*, while reducing her disappointment by γ. Thus, with our specification, the value of the outside option matches the customer's self-blame if, after choosing *In*, she followed the advisor's *Lie*. Put differently, the customer becomes more responsible for her low outcome the less costly is the alternative to trusting the advisor.

Only the remaining part of D_c (i.e., the difference between D_c and self-blame) is then attributed to the advisor, which is referred to as the advisor's "*shared guilt*". Thus, we effectively assume that the advisor shifts the causal responsibility for the outcome to the customer to the highest possible extent, by ignoring the part of the customer's disappointment for which the customer's own choice is pivotal.[13] Again, we first

[13] A more general specification of shared guilt, allowing for partial erosion of the advisor's responsibility, is considered in Appendix A. The qualitative theoretical predictions of the model still hold with this generalization.

define shared guilt for general strategy choices (s_c, s_a):

$$\hat{G}_a(s_a) = D_c(s_c, s_a) - G_c(s_c, s_a)$$

$$= D_c(s_c, s_a) - (D_c(s_c, s_a) - \min_{\tilde{s}_c} D_c(\tilde{s_c}, s_a)) = \min_{\tilde{s}_c} D_c(\tilde{s_c}, s_a)\,.^{14} \quad (12)$$

With this modification, the utility of the advisor is, as in the case of simple guilt (equation (4)),

$$U_a(s_c, s_a) = \pi_a(s_c, s_a) - \hat{\theta}_a \hat{G}_a(s_a), \quad (13)$$

where $\hat{\theta}_a$ is the sensitivity parameter with shared guilt. We now substitute for the customer's choice of *In* and the advisor's choice of *Lie*. Then the advisor's shared guilt is

$$\hat{G}_a(Lie) = \min_{\tilde{s}_c} D_c(\tilde{s_c}, Lie) = \alpha_c - \gamma, \quad (14)$$

which is indeed positive whenever the customer prefers to play *In* (so that $\alpha_c \geq \gamma$). Hence, with the considered choices of *In* and *Lie*, the advisor's expected utility depends again on his second-order belief about α_c, conditional on having observed *In* (and thus knowing that $\alpha_c \geq \gamma$). As the advisor's utility is linear in α_c from (13) and (14), the advisor's expected utility depends only on the respective first moment (of his updated beliefs), which we denoted by β_a^I. We thus have

$$U_a(Lie) = \pi_a^H - \hat{\theta}_a(\beta_a^I - \gamma), \quad (15)$$

where $\beta_a^I > \gamma$ by the definition in (6). Comparing this to the advisor's utility π_a^L from *Truth*, we now have that the advisor prefers lying to truth-telling if and only if

$$\pi_a^H - \hat{\theta}_a(\beta_a^I - \gamma) \geq \pi_a^L, \quad (16)$$

which is equivalent to

$$\hat{\theta}_a \leq \hat{\theta}_a^* = \frac{\pi_a^H - \pi_a^L}{\beta_a^I - \gamma}. \quad (17)$$

Importantly, in contrast to the cutoff obtained with simple guilt (see (9)), $\hat{\theta}_a^*$ depends now *directly* on γ for given β_a^I: (17) implies that, ceteris paribus, a higher γ increases the rate of lying. Once the customer foregoes a better outside option and still refers to the advisor, this action alleviates the advisor's (shared) guilt from lying. This is a novel, direct effect of γ on the advisor's strategy. At the same time, we still have an indirect effect that works through the advisor's second-order belief. Precisely, the cutoff still decreases with β_a^I as in the case of simple guilt, which in turn should be positively updated with γ; cf. equation (6).

[14] If the game involved additional players or chance moves, a more general model would add another layer of responsibility by also subtracting from $\hat{G}_a(s_a)$ the part of the customer's disappointment that *neither* the customer *nor* the advisor could avoid.

Thus, with shared guilt the effect of γ on the rate of lying $\hat{\theta}_a^*$ can be decomposed into two effects, where the first one is positive (guilt sharing) and the second one is negative (second-order belief induction). From (17), the overall effect depends on how the term $\beta_a^I - \gamma$ changes.

Prediction 2 [Shared Guilt]: *Under shared guilt, as the customer's foregone outside option increases:*

i) Keeping constant the advisor's belief about the trust of the customer, the advisor is more likely to lie.

ii) Taking into account also the induced change in the advisor's second-order belief, the advisor is more likely to lie if the effect of this belief change is weaker than the direct effect of γ (i.e., if $\beta_a^I - \gamma$ decreases).

Both, simple and shared guilt predict a negative relationship between lying and the advisor's second-order belief. Yet, unlike simple guilt, shared guilt allows for a positive relationship between the value of the foregone outside option and the likelihood of lying. In the next sections, we present results from laboratory experiments where we tested these predictions.

3. Experiment 1: How does the outside option matter for trustworthiness?

In this section, we show how advisors respond to customers who forego outside options with different values. The next section presents a robustness check.

3.1. Experimental design

The experimental game again involved two players (the advisor and the customer) and consisted of the following stages:

(1) The computer randomly chose a natural number from 1 to 100 with each number being equally likely. The advisor got privately informed about the actual number.

(2) The customer decided whether to directly purchase the information about the actual number at cost *c* (*Out*) or to refer to the advisor instead (*In*).

(3) The advisor observed the choice of the customer. If the customer had decided to refer to the advisor, the advisor had to send a message to the customer about the actual number (of the form 'The number is *m*', where *m* is a natural number between 1 and 100).

(4) The customer made a guess about the actual number, and the payoffs were realized.

	Advisor	*Customer in case of* In	*Customer in case of* Out
The guess is correct	**14**	**14**	**14 – *c***
The guess is incorrect	**17**	**4**	**Max{0,4 – *c*}**

Fig. 2. Payoffs (€) in Experiment 1.

The final payoffs depended on whether the customer's guess was correct, i.e., matched the computer number (see Fig. 2). The choice and payoff structure is equivalent to the Sharing Guilt Game depicted in Fig. 1 with $\epsilon = 1/99$ (the probability to guess correctly after not following a false advisor's message) and $\gamma = 14 - c$ (with other payoffs being accordingly rescaled).[15] The cost of information *c* varied between 2, 4, 6 and 8, and was deducted from the customer's payoff in case of *Out*.[16]

For reasons described in the next subsection, we implemented an additional treatment, called *No Choice Treatment* (NCT), where stage (2) was omitted. That is, the customer did not have a choice between *In* and *Out* and thus always had to refer to the advisor. Accordingly, the treatments with a possibility to buy information are from now on summarized as *Choice Treatment* (CT).

The game was played for 25 rounds with random rematching of subjects. NCT was played in three randomly chosen rounds and CT in the remaining 22 rounds (to ensure enough observations where the customer does not buy information for each treatment). The cost of information was chosen randomly each round (the same for all pairs), with higher frequency for smaller levels in order to counteract the decline of observations conditional on *In* at lower levels of *c*.[17] The price of information was made public knowledge between the players at the beginning of a round. At the end of a round, each player observed his/her payoff, while the customer was also shown the actual number chosen by the computer. At the end of the experiment, one round was randomly chosen for payment.

[15] Because the advisor's action space in the experimental game is not binary as in our Sharing Guilt Game, lying strategies of the advisor in the experimental game may theoretically subsume partially informative strategies, such as sending a truthful number with an added noise term. However, we do not find evidence that subjects were able to coordinate on equilibria with such partially informative communication. In particular, no customer was ever able to guess correctly after receiving a false number. This suggests that interpreting the advisor's strategy *Lie* in the Sharing Guilt Game to be equivalent to lying in the experimental game is a reasonable approximation, and it is the one we pursue here.

[16] In case of an incorrect guess after purchasing the information (which was never the case in the actual experiment), the customer's payoff was set at 0 if the cost of information exceeded 4 (see Fig. 2).

[17] The within-subject design was used to better control for possible heterogeneity of subjects' preferences in the statistical analysis. As Khalmetski et al. (2015) showed, controlling for such heterogeneity might have crucial implications for the detection of the effect of belief-dependent preferences on prosocial behavior.

To control for subjects' first- and second-order beliefs, customers were asked to estimate the expected rate of truth-telling in the current round (among advisors who must make the corresponding decision), and advisors were asked to estimate the answer to this question of their currently matched customer. We also asked both advisors and customers about the expected rate at which customers choose *Out*. Beliefs were elicited after the players made all their decisions in the round, but before they could observe their round payoffs. In particular, subjects were asked about their beliefs after customers made the choice between *In* and *Out*, advisors observed the respective customer's choice and send a message to her (in case of *In*), and customers made their guesses. Beliefs were measured in percentage points. We incentivized subjects by paying them additional four euros at the end of the experiment if the corresponding round was chosen for payment and if the actual value (i.e., the actual rate of truth-telling for customers and the matched customer's belief for advisors for the first question, and the actual rate of *Out* for the second question) did not differ from their guess by more than five percentage points. Players did not receive information about whether their beliefs earned an additional payment until the end of the experiment.

The experiment was conducted in the Cologne Laboratory for Economic Research with 236 participants split into eight sessions. Subjects were recruited with ORSEE (Greiner 2015), and the experiment was computerized with z-Tree (Fischbacher 2007). The average earning including a show-up fee of 2.5 euros was nearly 16 euros, and the experiment lasted for around 1.5 hours.[18] Translated instructions can be found in Online Appendix C.

3.2. Hypotheses

Simple guilt unambiguously predicts the following hypothesis for Experiment 1 (see Prediction 1).

Hypothesis 1 (Simple Guilt). *As the cost c of purchasing information becomes smaller, the advisor becomes more trustworthy.*

Shared guilt is consistent with Hypothesis 1, but additionally predicts that the customer's choice to forego a more attractive outside option leads to a shift in responsibility for her *ex post* disappointment towards herself. As a result, shared guilt can also result in more lying as *c* decreases (if the update in the advisor's second-order beliefs is not too intense, see Prediction 2). This leads to our competing Hypothesis 2.

Hypothesis 2 (Shared Guilt). *As the cost c of purchasing information becomes smaller, the advisor becomes less trustworthy.*

[18] After the experiment, the participants had to complete a questionnaire eliciting their age, gender, psychological measures of trust, self-assessment of understanding of the experimental instructions and interest in the experiment. Control questions regarding the participants' basic understanding of the instructions were also asked prior to the experiment.

Table 1. Comparison of theoretical predictions between different models.

	Simple guilt	*Intention-based reciprocity*	*Shared guilt*
Correlation between c and lying rate in CT	positive	negative	negative (controlling for β_a^I)
Relation between lying rates in CT and NCT	$CT < NCT$	$CT < NCT$	$CT > NCT$

Hence, a negative relationship between *c* and the rate of lying is consistent with shared guilt but not with simple guilt. Both theories, however, predict the same negative relationship, *ceteris paribus*, between the advisor's second-order belief and the rate of lying.

Intention-based reciprocity appears unlikely to play a role in our context as the customer's choice of *In* can never harm the advisor, and it can benefit the latter, and so can be interpreted as "kind", only if the customer is then deceived by the advisor. This outcome, however, is unlikely to be interpreted as the actual "intention" of the customer. Still, intention-based reciprocity models (such as Dufwenberg and Kirchsteiger 2004) can *in principle* be consistent with our Hypothesis 2. The idea is that a higher customer's expectation of truth-telling signaled by foregoing a better outside option can lead to a lower estimation of her kind intentions (associated with *In*) by the advisor, because the intended outcome would imply a lower advisor's payoff. This may, in reciprocal response, lead to a higher rate of lying by the advisor. If this is the case, predicted behavior would be as with shared guilt. To discriminate between the predictions of our model and those of intention-based reciprocity models, we added NCT to our experiment. The reciprocity models would predict that the rate of lying should be the highest in NCT, because – unlike in CT – there is no scope for positive reciprocity in NCT as there is no first mover's action that can be reciprocated. Shared guilt predicts the opposite effect: since the customer makes no choice (and hence cannot be held responsible for the outcome), there is no scope for guilt sharing so that the advisor's rate of lying should be the lowest in this treatment. This way, NCT can separate potential explanations based on the intention-based reciprocity and the shared-guilt model.

Simple guilt would also predict that the rate of lying in NCT is the highest among the experimental treatments, as in the case of reciprocity, because the second-order beliefs are not induced by the customer's choice of *In* and hence stay at the lowest level among all the treatments. At the same time, in contrast to reciprocity, simple guilt implies a positive correlation between *c* and the rate of lying in CT. Thus, our treatment variation allows us to distinguish between three competing theoretical explanations using data on the correlation of *c* with the lying rate, on the one hand, and the relation of the lying rate between CT and NCT, on the other hand, as summarized by Table 1.

Table 2. Rate of choosing *In* and conditional beliefs in Experiment 1, %.

Treatment		Customers' rate of choosing *In*	Customers' first-order beliefs		Advisors' second-order beliefs	
			conditional on *In*	conditional on *Out*	conditional on *In*	conditional on *Out*
CT	$c = 2$	7.9	62.7	38.2	47.6	32.5
	$c = 4$	20.2	61.0	36.5	47.7	32.4
	$c = 6$	49.4	52.2	30.8	43.3	30.5
	$c = 8$	65.8	44.1	31.9	40.5	27.4
NCT		-	40.6	-	38.9	-

First- and second-order beliefs measure the expected rate of truth-telling.

Other models that we are aware of, such as outcome-based models, are invariant to changes in *c*. The same applies to models that assume that truth-telling is driven by a fixed cost of lying, such as in Kartik (2009). Similarly, *c* plays no role in CT according to action-based reciprocity models (such as Cox et al. 2008, as opposed to intention-based models) once the customer's choice is locked in. No effect of *c* on advisor behavior is our null hypothesis.

3.3. Results

Table 2 presents the customer's rate of choosing *In* and the conditional first- and second-order beliefs.[19] Consistent with (1), the rate of *In* is increasing as the cost of choosing the outside option c increases.[20] Also in line with (1), customers who chose *In* at lower values of *c* are characterized by higher beliefs about the rate of advisors' truth-telling. In particular, beliefs conditional on *In* in CT($c = 2$) are significantly higher than in CT($c = 8$) as well as in NCT ($p = 0.009$ in both cases, one-sided Wilcoxon signed-rank test).[21] Those customers who chose *Out* have substantially lower beliefs about trustworthiness of advisors at all values of *c* (all pairwise differences are statistically significant at the 5%-level by one-sided Wilcoxon

[19] When reporting means here and below, the data is first aggregated at the matching group level (unless specified otherwise). This level of aggregation is also used for the non-parametric tests.

[20] The beliefs about the rate of *In* elicited from both customers and advisors are shown in Table B.1 in Online Appendix B.

[21] Here and below, we use one-sided tests whenever we have a clear directional hypothesis following from the model. The customers' first-order beliefs in 80.1% of cases are consistent with their choice in the sense of satisfying the incentive constraint (1).

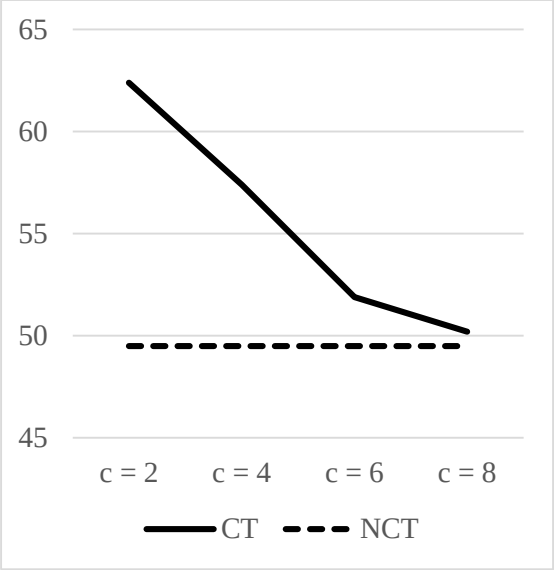

Fig. 3. Lying rate (%) in Experiment 1.

signed-rank test).[22] Advisors' second-order beliefs conditional on both *In* and *Out* follow a similar pattern, though the trend is less pronounced.[23] Yet, at the individual level, second-order beliefs conditional on *In* are strongly negatively correlated with the value of the outside option (Pearson correlation coefficient is -0.26, $p < 0.001$).[24] Thus, fully in line with the underlying mechanisms of both simple and shared guilt, a higher foregone outside option signals higher trust of the customer to advisors. Finally, after getting the message from the advisor, customers set their guess equal to the message in more than 95% of the cases at each level of c in CT and in more than 90% of the cases in NCT.

The main question is whether simple guilt correctly captures the advisors' responses to higher trust – namely more trustworthiness – or whether our refined model of shared guilt is a better predictor of advisor behavior. Fig. 3 shows that the lying rate in CT *decreases* as the foregone outside option of the customer gets worse, eventually converging to the lying rate in NCT. This is inconsistent with simple guilt, Hypothesis 1, but in line with shared guilt, Hypothesis 2. The result suggests that the effect of guilt sharing resulting from the customer foregoing a better outside option overcompensates the positive effect on prosocial behavior arising from signaling trust.

Regression analyses strongly confirm the overall picture. Table 3 presents the results of the random-effects probit model estimating the determinants of the lying rate in CT. Column 1 shows the basic

[22] The slight decrease of the latter beliefs with c is in line with the fact that, according to (1), it requires a larger degree of pessimism about the advisor's trustworthiness to choose *Out* at higher values of c.
[23] The difference in advisors' second-order beliefs conditional on *In* between CT at $c = 2$ and NCT is only marginally significant ($p = 0.062$, one-sided Wilcoxon signed-rank test).
[24] The estimation is based on the code provided in Stoddard (2011) and uses the data aggregated at the level of individual advisors in each treatment. An adjustment for repeated measurements within each subject is applied (Bland and Altman 1995).

Table 3. Determinants of the lying rate in Experiment 1 (CT), random-effects probit.

	(1)	(2)	(3)
Cost of information (c)	− 0.188***	− 0.249***	− 0.193***
	(0.029)	(0.029)	(0.062)
Round	0.060***	0.056***	0.089**
	(0.010)	(0.014)	(0.035)
Second-order belief		− 0.035***	− 0.034***
		(0.005)	(0.005)
Round × *c*			− 0.006
			(0.005)
Constant	0.593	2.489***	2.174***
	(0.483)	(0.471)	(0.589)
Observations	678	678	678

Second-order beliefs measure the expected rate of truth-telling. Standard errors (clustered at the matching group level) in parentheses.
** $p < 5\%$; *** $p < 1\%$.

specification with the exogenous experimental parameters as independent variables. The coefficient on the cost of information c is negative and highly significant, strongly supporting the predictions made by guilt sharing. The results of the non-parametric analysis, conducted for robustness, are also in line with shared guilt: The rate of lying at $c = 2$ is significantly higher than at $c = 8$ ($p = 0.042$, two-sided Wilcoxon signed-rank test).

Both simple and shared guilt predict a negative relationship between lying and the advisor's second-order belief due to the advisor's aversion to disappoint the customer's expectations. This is at the core of the baseline guilt aversion mechanism underlying both models. The regression analysis supports this assumption: The specification in column 2 adds advisors' second-order beliefs conditional on *In* to the analysis (β_a^I in our model), which were elicited after the advisor observed the customer's choice between *In* and *Out* in the current round. The effect of these beliefs on the rate of lying is shown to be significantly negative, in line with both the simple- and shared-guilt models.

The advisor's optimal choice in the model of simple guilt (as described in (9)) predicts no effect of the outside option γ once second-order beliefs are controlled for. That is, the total effect of the outside option should then be completely captured by the coefficient on second-order beliefs. In contrast, the model of shared guilt predicts a direct positive effect of γ on the rate of lying for a *given* level of the second-order belief, due to the change in the share of the customer's disappointment attributed exclusively to the advisor's action (see (17)). Our data supports shared guilt: The effect of the cost of information remains

highly negatively significant after controlling for second-order beliefs in column 2 of Table 3.[25] Ultimately, the guilt-sharing effect appears to be sufficiently strong to overcompensate potential second-order belief induction at higher outside options, so that the total effect of the cost of information *c* on the rate of lying is negative (column 1 of Table 3).

The positive relation between the foregone outside option and the resulting lying rate cannot be fully explained by intention-based reciprocity. As explained before, reciprocity presumes that the receiver considers the sender's choice of *In* as kind, so that the rate of lying in NCT (where no such choice is possible) would be higher than at any value of *c* in CT. Yet, we observe the opposite: the rate of lying is lower in NCT compared to any CT treatment. In particular, the rate of lying in NCT is 13% points lower than at CT with $c = 2$. At the same time, we note that the difference is only weakly statistically significant ($p = 0.069$, two-sided Wilcoxon signed-rank test), suggesting that some influence of intention-based reciprocity is not excluded by our data.

One important question is whether customers adjust their expectations across the treatments in the direction consistent with shared guilt, i.e., whether they anticipate that advisors become less trustworthy at lower values of *c*. Table 2 does not give an answer to this question since the variation of beliefs between the treatments there is affected by different patterns of customers' self-selection into *In* and *Out* at different values of *c*; for instance, more optimistic customers tend to be selected into *In* by lower values of *c*. In other words, this variation is affected by both between- and within-subject heterogeneity in expectations. To control for the between-subject component, Table 4 compares average first-order beliefs at different levels of $c < 8$ with the average first-order belief at $c = 8$ for *the same subsample of customers* (conditional on either *In* or *Out*).[26] The table shows, for instance, that while the average first-order belief of those customers who chose *In* at $c = 2$ is 58.6%,[27] *the very same* customers reported 42.1% on average when playing *In* at $c = 8$. More generally, the data shows that customers playing *In* tend to believe that advisors become more trustworthy when they observe *In* at a better outside option. Yet this change of belief was the opposite of the actual change of behavior of advisors in our experiment. All of the pairwise differences in beliefs are significant at the 5% level for customers choosing *In* (two-sided Wilcoxon signed-rank test). However, the

[25] Our regression results are robust to possible measurement errors regarding the elicitation of second-order beliefs. In particular, the effect of beliefs itself would be attenuated by the measurement error, going against our finding that second-order beliefs significantly affect lying. At the same time, under negative correlations between second-order beliefs and *c* and between second-order beliefs and the rate of lying, an error in measuring beliefs would imply an upward shift in the estimated coefficient on *c*, and hence would also go against the significantly negative coefficient on *c* which we obtain (see Gillen et al. 2015 for an illustrative example).

[26] The data is first aggregated at the subject level and then at the matching group level, to ensure equal weight for each subject (since individual observation frequencies are potentially endogenous in the considered case). The results are similar if the aggregation is performed directly at the matching group level.

[27] The difference to 62.7% reported in the previous Table 2 for the same case is due to the fact that here we have to restrict the subsample of customers to those who played *In* at least once at *both* $c = 2$ and $c = 8$.

Table 4. Within-subject variation of customers' first-order beliefs between treatments, %.

		First-order belief while playing *In* at $c = 8$	First-order belief while playing *In* at $c = X$
Subsample of customers who played *In* at least once both at $c = 8$ and $c = X$	$X = 2$	42.1	58.6
	$X = 4$	48.6	59.4
	$X = 6$	45.0	52.4
		First-order belief while playing *Out* at $c = 8$	First-order belief while playing *Out* at $c = X$
Subsample of customers who played *Out* at least once both at $c = 8$ and $c = X$	$X = 2$	31.3	35.7
	$X = 4$	31.3	35.8
	$X = 6$	30.3	33.0

First-order beliefs measure the expected rate of truth-telling.

effect is much smaller in magnitude when customers choose *Out*, being statistically significant only for customers choosing *Out* at $c = 2$ and $c = 6$.

This suggests that customers choosing *In* are particularly naïve in their mistaken belief that advisors react positively to the higher trust implied by larger foregone outside options. This naiveté is not inconsistent with our theoretical predictions, which do not rely on consistency of beliefs with behavior (see section 2.2.1). Such out-of-equilibrium beliefs tend to rather reinforce our main conclusion: Because customers do not anticipate the adverse effect of reducing the costs of information on the advisors' trustworthiness, they are even more vulnerable to these backfiring effects. However, as Table B.3 in Online Appendix B shows, this 'naïve' increase in expectations for decreasing *c* tends to diminish over time in many instances, which suggests that experience can be helpful in eliminating inaccurate beliefs.[28]

We finally take a look at how advisors' behavior changes over time, and whether our main effect vanishes with experience. Table 3 shows that there are more lies as advisors become more experienced (the *Round* effect is positive). Fig. 4 shows the lying rate over the tertiles of the game split over the experimental treatments. Advisors tend to lie more over time in all treatments. This complies with the results of Gneezy et al. (2013), who suggest that the phenomenon may be linked to the depletion of self-control. At the same time, we cannot exclude that this trend might also be affected by advisors learning their selfish strategy over time. However, Fig. 4 also shows that the trend is more pronounced for $c = 2$ and $c = 4$.[29] As a result, there is a larger spread of the lying rate between the treatments towards the last tertile of the game. Thus,

[28] This is a conservative estimate since many customers also switch from *In* to *Out* over time (see Table B.2 in Online Appendix B). Thus, Table B.3 shows that *even those* customers who keep to play *In* in later rounds tend to not expect anymore that advisors get more trustworthy with lower *c*.

[29] In line with this observation, the interaction term between the round number and *c* is negative (see column 3 in Table 3), and significantly so if by counting rounds one considers only the rounds where the advisor had to take the decision (see Table B.4 in Online Appendix B).

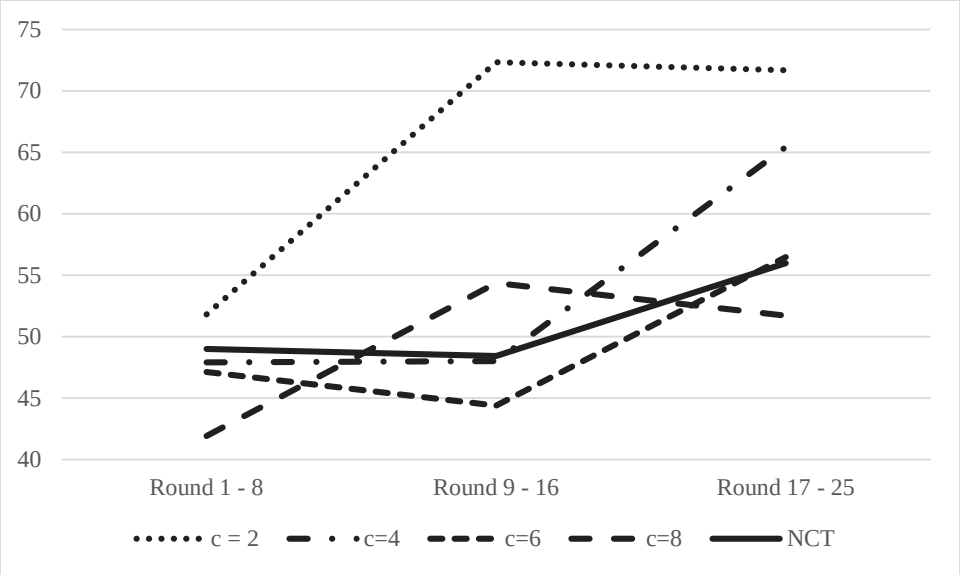

Fig. 4. Lying rate (%) over time in Experiment 1.

the inclination of advisors to lie more at lower values of *c* gets even stronger towards the end of the game, which suggests that this effect does not diminish with time and hence might be quite stable even in repeated interactions.

4. Experiment 2: Can self-selection explain our finding?

The main result of Experiment 1 is that in spite of the fact that the customer is supposed to be more trusting conditional on foregoing a better outside option, advisors become more likely to lie. Shared guilt suggests that this occurs due to a shift in the attribution of responsibility. As discussed in the introduction, however, the effect could be confounded by signaling or self-selection effects induced by the customer's choice of *In* conditional on a given outside option. For instance, asking the advisor for advice even when there would have been a cheap alternative to become informed might signal specific individual characteristics of a given customer, e.g., her greediness, or low risk/loss aversion, which in turn could possibly affect advisors' behavior towards this customer in the same direction as shared guilt. Because this cannot be excluded *ex ante*, we conducted Experiment 2, which allows to disentangle many potential selection and signaling effects of the customer's choice from the effect of a responsibility shift predicted by shared guilt.

4.1. Experimental design

The experimental game is the same as before, and is played repeatedly for 26 rounds with random rematching of players. Also as before, we had rounds with no customer's option to buy information, and

rounds that included such an option, with the same possible values of *c*. Yet, this time the treatments were played in a new, specific order. A round with no option to buy information always alternated with a round with such an option. At the beginning of each round with *no option* to buy information the advisor was informed about the choice of the currently matched customer in the *previous* round (while being reminded about the cost of information in that round). This is why we denote the rounds with no option to buy information as Previous Choice Treatment (PCT). The rounds *with an option* to buy information are denoted as Current Choice Treatment (CCT). CCT in Experiment 2 is equivalent to CT in Experiment 1 in terms of both (material) game and information structure, thus serving as a control treatment.

In each session, there were 13 rounds in CCT and 13 rounds in PCT. No subject played with the same partner in two consecutive rounds, which was made public knowledge. This ensured that the customer's choice in the previous round may affect the current advisor's behavior only via signaling of information about preferences/beliefs of the customer. The experiment was conducted in the Cologne Laboratory for Economic Research with 136 participants, split into 5 sessions. In each session subjects were divided into 2 independent matching groups. Translated instructions can be found in Online Appendix C.

4.2. Hypotheses

As in CT, the guilt-sharing model predicts an increase in the lying rate with the customer's foregone outside option in CCT. On the other side, PCT leaves no scope for self-blame of the customer (in the sense of (10)), as there is no current option to buy information, and hence the customer cannot retrospectively reduce her disappointment in case of lying by taking another choice.[30]

At the same time, all potential signaling effects of the customer's observed choice in PCT remain the same as in CCT. In particular, if the advisor reacts to the customer's (fixed) preferences as revealed by her behavior, it should generally not matter whether this information stems from her behavior in the current or in the previous round. Therefore, if the previously established positive correlation between the customer's unchosen outside option and the conditional lying rate was driven by some kind of signaling or selection effect, it should be observed in both PCT and CCT. At the same time, if this correlation was rather caused by the sharing of guilt induced by the (current) customer's choice of *In*, then this effect should not be observed in PCT while it should arise in CCT.[31]

[30] The customer can only slightly increase her expected payoff from 4 to 4.1 by not following the false message, as the probability to guess correctly at random is 1/99. This still leaves almost all of her disappointment attributed to the advisor according to (12).

[31] Still, one can argue that signaling effects induced by the customer's choice might be larger if taking place in the *current* interaction with the advisor, where the customer's choice becomes an integral part of this interaction. In other words, it is the customer's choice toward the relevant advisor (and not to any other) that changes the advisor's preferences towards the customer, and hence amends the subsequent response. If this is the case, our experimental design cannot completely rule out all signaling effects related to the customer's choice.

Table 5. Determinants of the lying rate in Experiment 2, random-effects probit.

	CCT		*PCT*	
	(1)	(2)	(3)	(4)
Cost of information (c)	− 0.199**	− 0.209**	0.038	0.046
	(0.100)	(0.091)	(0.057)	(0.069)
Round	0.057***	0.053**	0.037	0.035
	(0.015)	(0.024)	(0.030)	(0.025)
Second-order belief		− 0.038***		− 0.039***
		(0.008)		(0.007)
Constant	− 0.509	1.802**	− 1.301	0.814
	(0.578)	(0.850)	(0.887)	(0.889)
Observations	277	277	277	277

Second-order beliefs measure the expected rate of truth-telling. Standard errors (clustered at the matching group level) in parentheses.
** $p < 5\%$; *** $p < 1\%$.

4.3. Results

The results for the lying rate in CCT replicate the previously established effect of guilt sharing, as manifested by the regression results in Table 5 (columns 1 and 2). In particular, the coefficient on the cost of information *c* is significantly negative both with and without control for second-order beliefs. In stark contrast, the customer's outside option, foregone in the previous round, has no significant effect on the lying rate in PCT, with the effect, if at all, pointing even in the opposite direction (Table 5, columns 3 and 4).[32]

The results for the customer's behavior in CCT largely follow the observations in CT in Experiment 1 (Table 6). The rate of *In* increases with the cost of information, while the first-order beliefs of those customers who choose *In* decrease.[33] Advisors' second-order beliefs conditional on *In* also tend to follow a decreasing pattern with respect to *c* in PCT, and to a lesser extent in CCT.

We conclude that the previously established positive effect of the customer's unchosen outside option

[32] According to one-sided Wilcoxon signed-rank test (as we have a directional hypothesis based on Experiment 1), the rate of lying in CCT is significantly higher at $c = 2$ than at $c = 8$ with $p = 0.028$, being equal to 49.8% and 30.6% respectively. The rate of lying in PCT does not significantly differ between these two cost levels ($p = 0.572$), with the effect pointing even in the opposite direction relative to CCT: the aggregate rate of lying in PCT was 35.1% at $c = 2$ and 46.8% at $c = 8$.

[33] The beliefs about the rates of *In* elicited from both customers and advisors are shown in Table B.1 in Online Appendix B.

Table 6. Rate of choosing *In* and conditional beliefs in Experiment 2, %

Cost of information	Customers' rate of choosing *In* in CCT	Customers' first-order beliefs in CCT[34]		Advisors' second-order beliefs			
		conditional on *In*	conditional on *Out*	conditional on *In*		conditional on *Out*	
				CCT	*PCT*	*CCT*	*PCT*
$c = 2$	12.5	75.8	43.6	54.2	71.0	41.9	41.7
$c = 4$	31.3	66.1	44.1	55.4	52.8	40.2	43.8
$c = 6$	49.3	58.3	30.4	45.7	53.5	38.1	41.8
$c = 8$	69.6	58.4	33.6	50.6	48.6	44.3	41.5

First- and second-order beliefs measure the expected rate of truth-telling.

on the rate of lying was not driven by straightforward signaling or selection effects. The effect reveals itself only if the customer makes an explicit (trusting) choice in the current interaction, which, as our theory suggests, allows to reallocate responsibility for the final outcome to the customer.

5. Discussion and conclusion

In our game of strategic communication, advisors can assign a part of their responsibility for the final outcome to customers who, in turn, could have avoided being deceived by acquiring information on their own. Making the customers' information acquisition less costly led to advisors acting in a less trustworthy way. The effect cannot be explained by standard models of reciprocity, belief- or outcome-based preferences.

We propose a new model of shared causal attribution of guilt to explain our data. The model is a parsimonious and straightforward extension of a standard guilt aversion model (Battigalli and Dufwenberg 2007) often used in the relevant literature on trust and communication games, and we find significant support for this model's most basic predictions in our data, too. But only with our extension does the model capture our main findings. The model's innovation – sharing guilt – does not only appear to be intuitively plausible, but is also rooted in results in psychology and economics that robustly link the attribution of personal responsibility to the pivotality of one's behavior for the outcome. Taken together, we believe that these features make our approach a natural starting point to explain how reduced information costs to the customer can crowd out the advisor's sense of social responsibility.

[34] The customers' first-order beliefs in PCT constitute on average 50.0%.

Our findings may have a number of practical applications. They suggest that as the information age makes it easier for people to inform themselves, say, about product quality, competitors' prices or seller trustworthiness, customers are considered more accountable for their choices by sellers, which may reduce the sellers' responsibility to help them make an informed choice. Moreover, while such policies may be implemented to protect naïve customers, our data suggests that naïve customers are most vulnerable to the backfiring effect, because they are likely to (self-)select into trusting advisors at lower information costs.[35] Overall, our results should suggest to company policymakers and consumer protection regulators, such as in Internet markets or in the market for financial advice, that there can be unintended negative consequences in reducing consumers' need to rely on human experts, which might weaken the latter's sense of responsibility to behave in a trustworthy manner with respect to those consumers who still seek their advice.

Further research should be done to accommodate more complex structural characteristics of advisory communications. For instance, an investigation into whether the effect of guilt sharing persists in the case of *partial* information acquisition on the side of the customer appears promising. Another relevant question is whether the effect is robust to reputational concerns of the advisor in repeated interactions. From a methodological perspective, whether our concept of shared guilt is applicable to a broader class of trust games, e.g., to the games analyzed by Charness and Rabin (2002), remains an open question.

Our framework of shared guilt suggests a new perspective on the effect of increased informational transparency in the economy: better access to information may backfire. It also shows theoretically and empirically how responsibility and blame can be shared in strategic interactions. Given its success in accounting for our and previous laboratory data, we hope that our framework will prove to be useful in developing further applied models involving communication and trust.

Acknowledgements

We thank the Editor, the Associate Editor, three anonymous referees, Martin Dufwenberg, Nicholas Epley and seminar and conference participants in Cologne, Gothenburg and Kreuzlingen for very helpful comments and suggestions. Financial support of the German Research Foundation (DFG) through the Research Unit "Design and Behavior" (FOR 1371) and the ERC (Grant No. 229921) is gratefully acknowledged.

[35] Ambuehl et al. 2017 also study the selection of naïve subjects into participating in a risky activity.

References

Ahmad, F., Hudak, P.L., Bercovitz, K., Hollenberg, E., Levinson, W. (2006): "Are Physicians Ready for Patients with Internet-Based Health Information?" *Journal of Medical Internet Research*, 8(3), e22.

Akerlof, G.A. (1970): "The Market for 'Lemons': Quality Uncertainty and the Market Mechanism," *Quarterly Journal of Economics*, 84(3), 488–500.

Alicke, M. D., Mandel, D. R., Hilton, D. J., Gerstenberg, T., Lagnado, D. A. (2015): "Causal Conceptions in Social Explanation and Moral Evaluation: A Historical Tour," *Perspectives on Psychological Science*, 10(6), 790–812.

Ambuehl, S., Ockenfels, A., Stewart, C. (2017): "For They Know Not What They Do: Selection through Incentives when Information is Costly," working paper.

Anagol, S., Kim, H.H. (2012): "The Impact of Shrouded Fees: Evidence from a Natural Experiment in the Indian Mutual Funds Market," *American Economic Review*, 102(1), 576–593.

Attanasi, G., Battigalli, P., Manzoni, E. (2015): "Incomplete-Information Models of Guilt Aversion in the Trust Game," *Management Science*, 62(3), 648–667.

Bakos, J.Y. (1997): "Reducing Buyer Search Costs: Implications for Electronic Marketplaces," *Management Science*, 43(12), 1676–1692.

Bartling, B., Fischbacher, U. (2012): "Shifting the Blame: On Delegation and Responsibility," *Review of Economic Studies*, 79, 67–87.

Bartling, B., Fischbacher, U., Schudy, S. (2015): "Pivotality and Responsibility Attribution in Sequential Voting," *Journal of Public Economics*, 128, 133–139.

Battigalli, P., Dufwenberg, M. (2007): "Guilt in Games," *American Economic Review*, 97, 170–176.

Battigalli, P., Dufwenberg, M. (2009): "Dynamic Psychological Games," *Journal of Economic Theory*, 141, 1–35.

Beck, A., Kerschbamer, R., Qiu, J., Sutter, M. (2013): "Shaping Beliefs in Experimental Markets for Expert Services: Guilt Aversion and the Impact of Promises and Money-Burning Options," *Games and Economic Behavior*, 81, 145–164.

Bellemare, C., Sebald, A., Strobel, M. (2011): "Measuring the Willingness to Pay to Avoid Guilt: Estimation Using Equilibrium and Stated Belief Models," *Journal of Applied Econometrics*, 26, 437–453.

Bland, J.M, Altman, D.G. (1995): "Calculating Correlation Coefficients with Repeated Observations: Part 2 - Correlation between Subjects," *BMJ*, 310, 633.

Bolton, G.E., Ockenfels, A. (2000): "ERC: A Theory of Equity, Reciprocity, and Competition," *American Economic Review*, 90, 166–193.

Bolton, G., Greiner, B., Ockenfels, A. (2013): "Engineering Trust: Reciprocity in the Production of Reputation Information," *Management Science*, 59(2), 265–285.

Brandts, J., Fatas, E., Haruvy, E., Lagos, F. (2015): "The Impact of Relative Position and Returns on Sacrifice and Reciprocity: An Experimental Study Using Individual Decisions," *Social Choice and Welfare*, 45(3), 489–511.

Cain, D.M., Loewenstein, G., Moore, D.A. (2005): "The Dirt on Coming Clean: Perverse Effects of Disclosing Conflicts of Interest," *Journal of Legal Studies*, 34(1), 1–25.

Cain, D.M., Loewenstein, G., Moore, D.A. (2011): "When Sunlight Fails to Disinfect: Understanding the Perverse Effects of Disclosing Conflicts of Interest," *Journal of Consumer Research*, 37(5), 836–857.

Cappelen, A.W., Fest, S., Sørensen, E.Ø., Tungodden, B. (2016): "Choice and Personal Responsibility: What Is a Morally Relevant Choice," Discussion Paper 27/2014, NHH Department of Economics.

Charness, G. (2000): "Responsibility and Effort in an Experimental Labor Market," *Journal of Economic Behavior and Organization*, 42(3), 375–384.

Charness, G., Cobo-Reyes, R., Jiménez, N., Lacomba, J.A., Lagos, F. (2012): "The Hidden Advantage of Delegation: Pareto Improvements in a Gift Exchange Game," *American Economic Review*, 102(5), 2358–2379.

Charness, G., Dufwenberg, M. (2006): "Promises and Partnership," *Econometrica*, 74, 1579–1601.

Charness, G., Rabin, M. (2002): "Understanding Social Preferences with Simple Tests," *Quarterly Journal of Economics*, 117(3), 817–869.

Coffman, L.C. (2011): "Intermediation Reduces Punishment (and Reward)," *American Economic Journal: Microeconomics*, 3(4), 77–106.

Cox, J.C., Friedman, D., Sadiraj, V. (2008): "Revealed Altruism," *Econometrica*, 76(1), 31–69.

Cox, J.C., Servátka, M., Vadovič, R. (2010): "Saliency of Outside Options in the Lost Wallet Game," *Experimental Economics*, 13, 66–74.

Davis, C.G., Lehman, D.R., Silver, R.C., Wortman, C.B., Ellard, J.H. (1996): "Self-Blame Following a Traumatic Event: The Role of Perceived Avoidability," *Personality and Social Psychology Bulletin*, 22, 557–567.

Duarte, F, Hastings, J.S. (2012): "Fettered Consumers and Sophisticated Firms: Evidence from Mexico's Privatized Social Security Market," working paper no. w18582, National Bureau of Economic Research.

Dufwenberg, M., Gneezy, U. (2000): "Measuring Beliefs in an Experimental Lost Wallet Game," *Games and Economic Behavior*, 30, 163–182.

Dufwenberg, M., Kirchsteiger, G. (2004): "A Theory of Sequential Reciprocity," *Games and Economic Behavior*, 47(2), 268–298.

Ellingsen, T., Johannesson, M. (2008): "Pride and Prejudice: The Human Side of Incentive Theory," *American Economic Review*, 98(3), 990–1008.

Ellingsen, T., Johannesson, M., Tjøtta, S., Torsvik, G. (2010): "Testing Guilt Aversion," *Games and Economic Behavior*, 68(1), 95–107.

Engl, F. (2017): "A Theory of Causal Responsibility Attribution," working paper, University of Cologne.

Erat, S., Gneezy, U. (2012): "White Lies," *Management Science*, 58(4), 723–733.

Falk, A., Szech, N. (2013): "Organizations, Diffused Pivotality and Immoral Outcomes," CESifo Working Paper Series No. 4300.

Fehr, E., Schmidt, K.M. (1999): "A Theory of Fairness, Competition and Cooperation," *Quarterly Journal of Economics*, 114, 817–868.

Fershtman, C., Gneezy, U. (2001): "Strategic Delegation: An Experiment," *RAND Journal of Economics*, 32(2), 352–368.

Fischbacher, U. (2007): "z-Tree: Zurich Toolbox For Ready-Made Economic Experiments," *Experimental Economics*, 10, 171−178.

Garofalo, O., Rott, C. “Shifting Blame? Experimental Evidence on Delegating Communication,” *Management Science*, forthcoming.

Geanakoplos, J., Pearce, D., Stacchetti, E. (1989): “Psychological Games and Sequential Rationality,” *Games and Economic Behavior*, 1, 60–79.

Gillen, B., Snowberg, E., Yariv, L. (2015): “Experimenting with Measurement Error: Techniques with Applications to the Caltech Cohort Study,” National Bureau of Economic Research (No. w21517).

Gneezy, U. (2005): “Deception: The Role of Consequences,” *American Economic Review*, 95, 384–394.

Gneezy, U., Rockenbach, B., Serra-Garcia, M. (2013): “Measuring Lying Aversion,” *Journal of Economic Behavior and Organization*, 93, 293–300.

Golan, E., Kuchler, F., Mitchell, L., Greene, C., Jessup, A. (2001): “Economics of Food Labeling,” *Journal of Consumer Policy*, 24(2), 117–184.

Goldmanis, M., Hortaçsu, A., Syverson, C., Emre, Ö. (2010): “E-Commerce and the Market Structure of Retail Industries,” *Economic Journal*, 120(545), 651–682.

Greiner, B. (2015): “Subject Pool Recruitment Procedures: Organizing Experiments with ORSEE,” *Journal of the Economic Science Association*, 1(1), 114–125.

Hamman, J.R., Loewenstein, G., Weber, R.A. (2010): “Self-Interest through Delegation: An Additional Rationale for The Principal-Agent Relationship,” *American Economic Review*, 100(4), 1826–1846.

Inderst, R., Ottaviani, M. (2012): “Competition through Commissions and Kickbacks,” *American Economic Review*, 102(2), 780–809.

Kartik, N. (2009): “Strategic Communication with Lying Costs,” *Review of Economic Studies*, 76, 1359–1395.

Kaufmann, D., Bellver, A. (2005): “Transparenting Transparency: Initial Empirics and Policy Applications,” working paper, World Bank.

Kawagoe, T., Narita, Y., (2014): “Guilt Aversion Revisited: An Experimental Test of a New Model,” *Journal of Economic Behavior and Organization*, 102, 1–9.

Khalmetski, K. (2016): “Testing Guilt Aversion with an Exogenous Shift in Beliefs,” *Games and Economic Behavior*, 97, 110–119.

Khalmetski, K. Ockenfels, A., Werner, P. (2015): "Surprising Gifts: Theory and Laboratory Evidence," *Journal of Economic Theory*, 159(A), 163–208.

Kozup, J., Hogarth, J.M. (2008): "Financial Literacy, Public Policy, and Consumers' Self-Protection – More Questions, Fewer Answers," *Journal of Consumer Affairs*, 42(2), 127–136.

Lacko, J., Pappalardo, J. (2007): "Improving Consumer Mortgage Disclosures: An Empirical Assessment of Current and Prototype Disclosure Forms," Federal Trade Commission Bureau of Economics Staff Report, Washington, DC.

Levine, D.K. (1998): "Modeling Altruism and Spitefulness in Experiments," *Review of Economic Dynamics*, 1(3), 593–622.

Mandel, D. (2003): "Counterfactuals, Emotions, and Context," *Cognition and Emotion*, 17(1), 139–159.

Miller, D.T., Turnbull, W. (1990): "The Counterfactual Fallacy: Confusing What Might Have Been with What Ought to Have Been," *Social Justice Research*, 4(1), 1–19.

Oexl, R., Grossman, Z.J. (2013): "Shifting the Blame to a Powerless Intermediary," *Experimental Economics*, 16(3), 306–312.

Reuben, E., Sapienza, P., Zingales, L. (2009): "Is Mistrust Self–Fulfilling?" *Economics Letters*, 104, 89–91.

Servátka, M., Vadovič, R. (2009): "Unequal Outside Options in the Lost Wallet Game," *Economics Bulletin*, 29(4), 2870–2883.

Stiglitz, J.E., Weiss, A. (1981): "Credit Rationing in Markets with Imperfect Information," *American Economic Review*, 71(3), 393–410.

Stoddard G.J. (2011): "Biostatistics and Epidemiology Using Stata: A Course Manual," unpublished manuscript, University of Utah School of Medicine.

Vanberg, C. (2008): "Why Do People Keep Their Promises? An Experimental Test of Two Explanations," *Econometrica*, 76(6), 1467–1480.

Woods, D., Servátka, M. (2016): Testing Psychological Forward Induction and the Updating of Beliefs in the Lost Wallet Game," *Journal of Economic Psychology*, 56, 116–125.

Appendix A: Generalized Model

We now derive Hypothesis 2 more formally. In doing so, we also generalize our results by allowing for more general preferences for the advisor. Specifically, in our main specification, we assumed that the advisor does not feel guilt for the share of the customer's disappointment that the customer could have avoided by her own action. We now relax this specification by allowing the advisor to feel responsible also for this share of the customer's disappointment to a certain extent, with the degree of pivotality of the customer's own action only partially mitigating the advisor's guilt.

We first specify two general terms:

$$\tau_a(s_c, s_a) = D_c(s_c, s_a) - \min_{\widetilde{s_a}} D_c(s_c, \widetilde{s_a}), \tag{18}$$

$$\tau_c(s_c, s_a) = D_c(s_c, s_a) - \min_{\widetilde{s_c}} D_c(\widetilde{s_c}, s_a). \tag{19}$$

Here, τ_a is the part of the customer's disappointment for which the advisor's choice is pivotal (i.e., a necessary condition), while τ_c denotes the part of the customer's disappointment for which the customer's own choice is pivotal. Note that τ_a is also equivalent to the advisor's simple guilt, while τ_c is equivalent to the customer's self-blame in terms of our previous definitions. The general guilt function of the advisor is then assumed to increase in the degree of pivotality of the advisor (in line with simple guilt), and decrease in the degree of pivotality of the customer (as a channel of responsibility shifting):

$$\tilde{G}_a(s_c, s_a) = f(\tau_a, \tau_c), \tag{20}$$

$$f_1(\tau_a, \tau_c) \geq 0, \tag{21}$$

$$f_2(\tau_a, \tau_c) \leq 0. \tag{22}$$

From (5) and (11) we obtain

$$\tau_a(In, Lie) = \alpha_c, \tag{23}$$

$$\tau_c(In, Lie) = \gamma, \tag{24}$$

which implies

$$\tilde{G}_a(In, Lie) = f(\alpha_c, \gamma). \tag{25}$$

As before, the guilt term is subtracted from the monetary utility so that the advisor's expected utility from lying after *In* is

$$U_a(In, Lie) = \pi_a^H - \tilde{\theta}_a E_a[f(\alpha_c, \gamma)]. \tag{26}$$

Comparing this to the advisor's utility π_a^L from choosing *Truth* gives rise to the cutoff

$$\tilde{\theta}_a^* = \frac{\pi_a^H - \pi_a^L}{E_a[f(\alpha_c, \gamma)]}. \tag{27}$$

Further, since the customer chooses *In* if and only if $\alpha_c \geq \gamma$, we have

$$\tilde{\theta}_a^* = \frac{\pi_a^H - \pi_a^L}{E_a[f(\alpha_c, \gamma)]} = \frac{\pi_a^H - \pi_a^L}{\frac{1}{1-H(\gamma)} \int_\gamma^1 f(\alpha_c, \gamma) h(\alpha_c) d\alpha_c}, \tag{28}$$

where h and H are, respectively, probability density and cumulative distribution functions of the customer's first-order beliefs as expected by the advisor.

Imposing now that the respective variables are all continuously differentiable, we obtain

$$\frac{d\tilde{\theta}_a^*}{d\gamma} = \frac{\pi_a^H - \pi_a^L}{\left(\int_\gamma^1 f(\alpha_c, \gamma) h(\alpha_c) d\alpha_c\right)^2} (\kappa_1 - \kappa_2), \tag{29}$$

where

$$\kappa_1 = -\big(1 - H(\gamma)\big) \int_\gamma^1 f_2(\alpha_c, \gamma) h(\alpha_c) d\alpha_c \geq 0, \tag{30}$$

$$\kappa_2 = h(\gamma) \int_\gamma^1 \big(f(\alpha_c, \gamma) - f(\gamma, \gamma)\big) h(\alpha_c) d\alpha_c \geq 0. \tag{31}$$

Note that κ_1 is increasing in the magnitude of $f_2(\alpha_c, \gamma)$ for given α_c, i.e., with the strength of responsibility shifting towards the customer as γ gets larger. The second term κ_2 rather captures the effect of induced guilt aversion due to the advisor having higher conditional second-order beliefs by higher levels of γ, which shifts the guilt function upwards. For instance, if most of the probability mass assigned to customers' beliefs in $[\gamma, 1]$ is put on beliefs just slightly above γ (where the term $f(\alpha_c, \gamma) - f(\gamma, \gamma)$ is small), the whole term κ_2 is also small.

By (29), $\tilde{\theta}_a^*$ (which determines the lying rate) increases with γ if

$$\kappa_1 > \kappa_2 \tag{32}$$

$$\Leftrightarrow -\big(1 - H(\gamma)\big) \int_\gamma^1 f_2(\alpha_c, \gamma) h(\alpha_c) d\alpha_c > h(\gamma) \int_\gamma^1 \big(f(\alpha_c, \gamma) - f(\gamma, \gamma)\big) h(\alpha_c) d\alpha_c. \tag{33}$$

Hence, there is a trade-off between two effects: guilt alleviation due to the shift of responsibility towards the customer (captured by the left-hand side of (33)), and the increase in the expected customer's disappointment due to self-selection of customers with higher expectations into *In* (captured by the right-

hand side of (33)). In particular, if the self-selection effect is small, then it is more likely that the first effect dominates, i.e., that the rate of lying increases with the value of the foregone outside option.[36]

Appendix B: Supplementary statistical analysis

Table B.1. Expected rate of choosing *In, %.*

	Experiment 1 (CT)		Experiment 2 (CCT)	
	Customers	Advisors	Customers	Advisors
$c = 2$	17.9	14.7	21.7	20.9
$c = 4$	27.5	24.5	30.8	36.1
$c = 6$	40.5	36.0	42.1	42.1
$c = 8$	53.0	48.2	52.2	60.1

The expected rates of *In* are converted from the reported expected rates of *Out.*

Table B.2. The rate of choosing *In* over experimental rounds in Experiment 1, %.

Treatment		Rounds		
		1-8	*9-16*	*17-25*
CT	$c = 2$	14.6	5.6	2.7
	$c = 4$	27.9	17.9	18.6
	$c = 6$	54.5	49.9	46.2
	$c = 8$	71.4	70.9	62.9
NCT		-	-	-

[36] A further layer of endogeneity can be added into the model by assuming that the customer's first-order belief is sensitive to the value of the outside option, i.e., that a given individual belief (and hence its distribution $h(\alpha_c)$) is a function of γ. This can further affect the direction in which the tradeoff specified in (33) is likely to be resolved.

Table B.3. Within-subject variation of customers' first-order beliefs between treatments in Experiment 1, over experimental rounds, %.

Rounds 1-8:			
		First-order belief while playing *In* at $c = 8$	First-order belief while playing *In* at $c = X$
Subsample of customers who played *In* at least once both at $c = 8$ and $c = X$	$X = 2$	68.0	72.1
	$X = 4$	61.3	71.7
	$X = 6$	48.4	60.8
		First-order belief while playing *Out* at $c = 8$	First-order belief while playing *Out* at $c = X$
Subsample of customers who played *Out* at least once both at $c = 8$ and $c = X$	$X = 2$	31.7	42.3
	$X = 4$	27.6	37.4
	$X = 6$	31.2	38.3
Rounds 9-16:			
		First-order belief while playing *In* at $c = 8$	First-order belief while playing *In* at $c = X$
Subsample of customers who played *In* at least once both at $c = 8$ and $c = X$	$X = 2$	67.8	73.4
	$X = 4$	73.1	75.7
	$X = 6$	53.6	55.6
		First-order belief while playing *Out* at $c = 8$	First-order belief while playing *Out* at $c = X$
Subsample of customers who played *Out* at least once both at $c = 8$ and $c = X$	$X = 2$	35.6	36.3
	$X = 4$	35.6	40.6
	$X = 6$	35.9	38.2
Rounds 17-25:			
		First-order belief while playing *In* at $c = 8$	First-order belief while playing *In* at $c = X$
Subsample of customers who played *In* at least once both at $c = 8$ and $c = X$	$X = 2$	32.3	36.3
	$X = 4$	45.8	47.0
	$X = 6$	45.3	45.9
		First-order belief while playing *Out* at $c = 8$	First-order belief while playing *Out* at $c = X$
Subsample of customers who played *Out* at least once both at $c = 8$ and $c = X$	$X = 2$	34.9	34.7
	$X = 4$	34.9	32.8
	$X = 6$	33.3	34.7

First-order beliefs measure the expected rate of truth-telling.

Table B.4. Determinants of the lying rate in Experiment 1 (CT), random-effects probit.

	(1)	(2)	(3)
Cost of information (c)	− 0.181***	− 0.251***	− 0.172***
	(0.033)	(0.033)	(0.047)
Active round	0.136***	0.143***	0.241***
	(0.034)	(0.032)	(0.054)
Second-order belief		− 0.036***	− 0.036***
		(0.005)	(0.005)
Active round × c			− 0.017**
			(0.007)
Constant	0.585	2.529***	2.070***
	(0.466)	(0.446)	(0.525)
Observations	678	678	678

Second-order beliefs measure the expected rate of truth-telling. Active round stands for the number of rounds, including the current one, where the advisor has had to take a decision. Standard errors (clustered at the matching group level) in parentheses.
** $p < 5\%$; *** $p < 1\%$

Appendix C: Experimental instructions (translated from German)

Experiment 1

Welcome to our experiment!

You are now participating in an experiment in which you can earn money. Your total payoff depends on your own decisions and those of the other participants. Please refrain from now on from talking and looking at other participants' screens. If you have any questions, please raise your hand. We will come to your place and answer your question as soon as possible. During the experiment you will interact with other participants. The identity of the other participants will not be revealed to you. Likewise, your identity will not be revealed to the other participants. These instructions are identical for all participants.

Before the experiment begins, you will be assigned the role of either the "advisor" or the "customer". This role will be retained during the entire experiment.

The experiment consists of 25 rounds. In every round, each advisor is matched to a new customer. The matching in every round is random.

In every round, the computer randomly draws a "secret" number between 1 and 100 for each participant pair. Each number has the same chance of being drawn. The advisor learns the drawn number at the beginning of each round but the customer does not.

The customer's task is to guess the "secret" number. The payoffs of both players depend on whether the customer's guess is correct or false, and are as follows (in "game points"):

	Points Advisor	Points Customer
The customer's guess is *correct*	**14**	**14**
The customer's guess is *false*	**17**	**4**

In some rounds, the customer can acquire information about the secret number only from his advisor. In the other rounds, the customer has an additional option to acquire information about the secret number by himself (without his advisor). However, this information is not free and costs the customer 2, 4, 6 or 8 points. Both participants are informed at the beginning of each round whether or not the customer can acquire information about the secret number and how much it costs. If the customer decides to acquire information, the actual secret number is revealed to him. The information costs are then subtracted from his payoff for the round. If the payoff is not sufficient to cover the information costs, his final payoff for the round is 0 points.

If the customer decides to proceed without buying information from the computer or if this option is not available, the advisor must send to the customer a message in the following form:

"The secret number is …"

The advisor can transmit any possible number independently of the actual number.

After the customer has bought information or has received a message from his advisor, he has to make a guess. At the very end of the round, each participant learns how many points he and his fellow player received in this round. Additionally, the customer learns the actual number that was observed by the advisor.

At the end of the entire experiment, one of the 25 rounds is randomly chosen for payment. The same round is chosen for all participants. *For the payout, 1 point equals to one euro.* Additionally, you always receive 2.50 euros for showing up.

After signing the receipt, you will receive your payoff in cash.

At the end of the experiment, we will also ask you to answer a short questionnaire on your computer.

Please click the button "Done" once you have read and understood the instructions.

Experiment 2

Welcome to our experiment!

You are now participating in an experiment in which you can earn money. Your total payoff depends on your own decisions and those of the other participants. Please refrain from now on from talking and looking at other participants' screens. If you have any questions, please raise your hand. We will come to your place and answer your question as soon as possible. During the experiment you will interact with other participants. The identity of the other participants will not be revealed to you. Likewise, your identity will not be revealed to the other participants. These instructions are identical for all participants.

Before the experiment begins, you will be assigned the role of either the "advisor" or the "customer". This role will be retained during the entire experiment.

The experiment consists of 26 rounds. In every round, each advisor is matched to a new customer. The matching in every round is random. The matching ensures that the same advisor and customer never interact with each other in two successive rounds.

In every round, the computer randomly draws a "secret" number between 1 and 100 for each participant pair. Each number has the same chance of being drawn. The advisor learns the drawn number at the beginning of each round but the customer does not.

The customer's task is to guess the "secret" number. The payoffs of both players depend on whether the customer's guess is correct or false, and are as follows (in "game points"):

	Points Advisor	Points Customer
The customer's guess is *correct*	**14**	**14**
The customer's guess is *false*	**17**	**4**

In some rounds, the customer can acquire information about the secret number only from his advisor. In the other rounds, the customer has an additional option to acquire information about the secret number by himself (without his advisor). However, this information is not free and costs the customer 2, 4, 6 or 8 points. Both participants are informed at the beginning of each round whether or not the customer can acquire information about the secret number and how much it costs. If the customer decides to acquire information, the actual secret number is revealed to him. The information costs are then subtracted from his payoff for the round. If the payoff is not sufficient to cover the information costs, his final payoff for the round is 0 points.

The advisor is immediately notified whether his customer has bought information. Additionally, the advisor learns whether the customer bought information in the previous round (while he was assigned to a different advisor).

If the customer decides to proceed without buying information from the computer or if this option is not available, the advisor must send to the customer a message in the following form:

"The secret number is …"

The advisor can transmit any possible number independently of the actual number.

After the customer has bought information or has received a message from his advisor, he has to make a guess. At the very end of the round, each participant learns how many points he and his fellow player received in this round. Additionally, the customer learns the actual number that was observed by the advisor.

At the end of the entire experiment, one of the 26 rounds is randomly chosen for payment. The same round is chosen for all participants. *For the payout, 1 point equals to one euro.* Additionally, you always receive 2.50 euros for showing up.

After signing the receipt, you will receive your payoff in cash.

At the end of the experiment, we will also ask you to answer a short questionnaire on your computer.

Please click the button "Done" once you have read and understood the instructions.

Belief elicitation questions (*shown on computer screens at the end of each round in both experiments*):

Control Question 1 [Customers]:

Reminder: In this round, the customers could not [could] acquire information themselves [for … points].[37]

We now ask you for your guess. What percentage of the advisors participating in this experiment and having had to send a message in this round have sent the true number to their customer?

If your guess does not deviate from the actual number by more than 5%, *you will get an additional bonus of 4 euros* (in case this round is relevant for your payment). If no advisor had to send a message in this round, you will get no additional payoff.

Please provide your guess (from 0 to 100) with at most one decimal point.

You will be informed at the end of the experiment if you have earned an additional payoff with your guess.

[37] For this and subsequent questions, the following sentence was preceding the text of the question in case it has been already asked before: "Please answer Control Question 1[2] (which is the same as before), but now with respect to *this round. To remind you: …*"

Control Question 1 [Advisors]:

Reminder: In this round, the customers could not [could] acquire information themselves [for … points].

We have asked your customer about his guess: "What percentage of the advisors participating in this experiment and having had to send a message in this round have sent the true number to their customer?" We now ask you to guess what answer your customer has given.

If your guess does not deviate from the actual number by more than 5%, *you will get an additional bonus of 4 euros* (in case this round is relevant for your payment).

Please provide your guess (from 0 to 100) with at most one decimal point.

You will be informed at the end of the experiment if you have earned an additional payoff with your guess.

Control Question 2 [Customers, only in CT/CCT]:

Reminder: In this round, the customers could not [could] acquire information themselves [for … points].

We now ask you for your guess. What percentage of the customers participating in this experiment (including yourself) have acquired information about the secret number from the computer *in this round*?

If your guess does not deviate from the actual number by more than 5%, *you will get an additional bonus of 4 euros* (in case this round is relevant for your payment).

Please provide your guess (from 0 to 100) with at most one decimal point.

You will be informed at the end of the experiment if you have earned an additional payoff with your guess.

Control Question 2 [Advisors, only in CT/CCT]:

Reminder: In this round the customers could not [could] acquire information themselves [for … points].

We now ask you for your guess. What percentage of the customers participating in this experiment have acquired information about the secret number from the computer *in this round*?

If your guess does not deviate from the actual number by more than 5%, *you will get an additional bonus of 4 euros* (in case this round is relevant for your payment).

Please provide your guess (from 0 to 100) with at most one decimal point.

You will be informed at the end of the experiment if you have earned an additional payoff with your guess.